Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images

Sachin Mehta[1], Ezgi Mercan[1], Jamen Bartlett[2], Donald Weaver[2], Joann Elmore[3], and Linda Shapiro[1]

[1]University of Washington, Seattle   [2]University of Vermont, Burlington   [3]University of California, Los Angeles

Email: {sacmehta, ezgi, shapiro}@cs.washington.edu {jamen.bartlett, donald.weaver}@uvmhealth.org elmore@mednet.ucla.edu

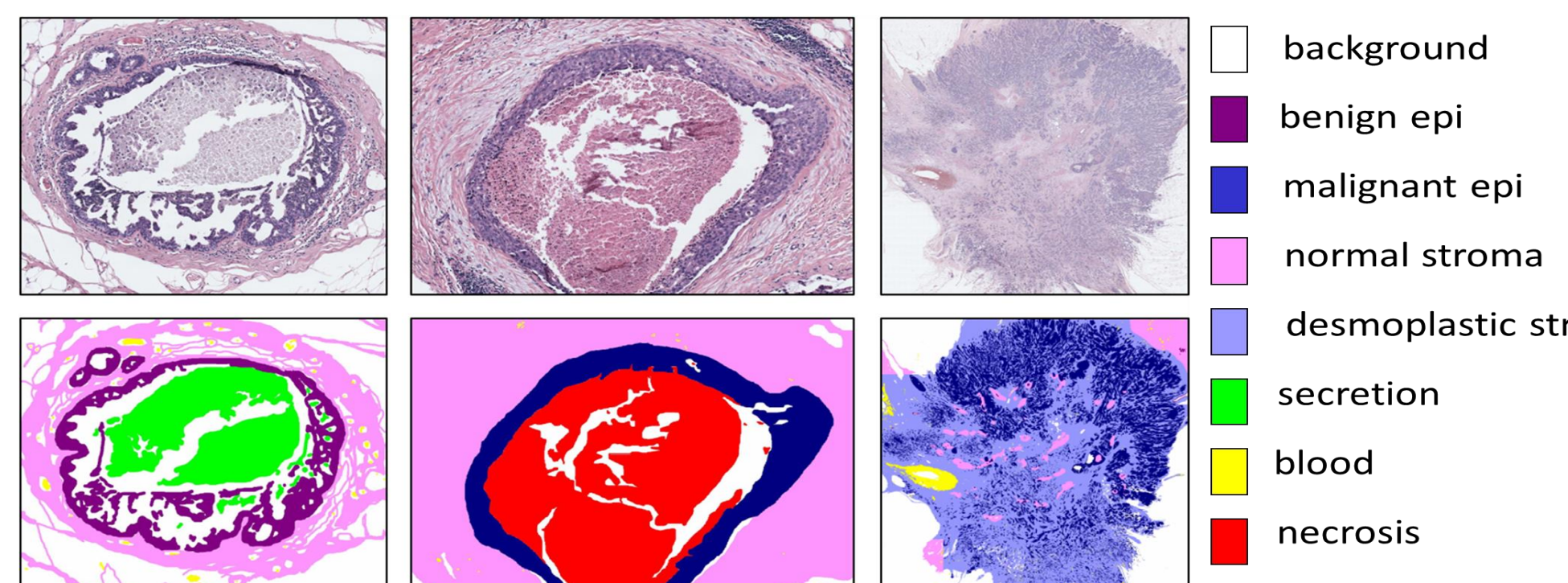**Project webpage**

MICCAI2018 Granada SPAIN

Introduction

- **One in eight** women in the United States will be diagnosed with breast cancer in their lifetime.
- Diagnostic errors are **alarmingly frequent**, lead to incorrect treatment recommendations, and can cause significant patient harm.
- Unlike standard image datasets, breast biopsy images have objects of interest in **varied sizes and shapes**.
- **Saliency-based methods** can identify regions of interest that could aid in diagnosis; however, they fail to provide structure- and tissue-level information.
- **Semantic segmentation-based methods** provide **a powerful abstraction** so that simple features with diagnostic classifiers, like multi-layer perceptron, perform well for **automated diagnosis**. However, these approaches cannot weigh the importance of different tissue types.
- We introduce **Y-Net** that combines these two independent approaches to generate **discriminative segmentation masks**.

Breast biopsy dataset

- Our dataset consists of *240 whole slide images (WSIs),* which are classified into 4 diagnostic categories (**benign, atypia, ductal carcinoma *in situ*, and invasive cancer**) by *87 pathologists*. Each slide was interpreted by a panel of three experts to assign a consensus diagnostic label.
- Pathologists also marked *428 region of interest* (ROI) that helped in diagnosis. A subset of *58 of these ROIs* have been hand segmented by a pathology fellow into **eight different tissue classifications**.
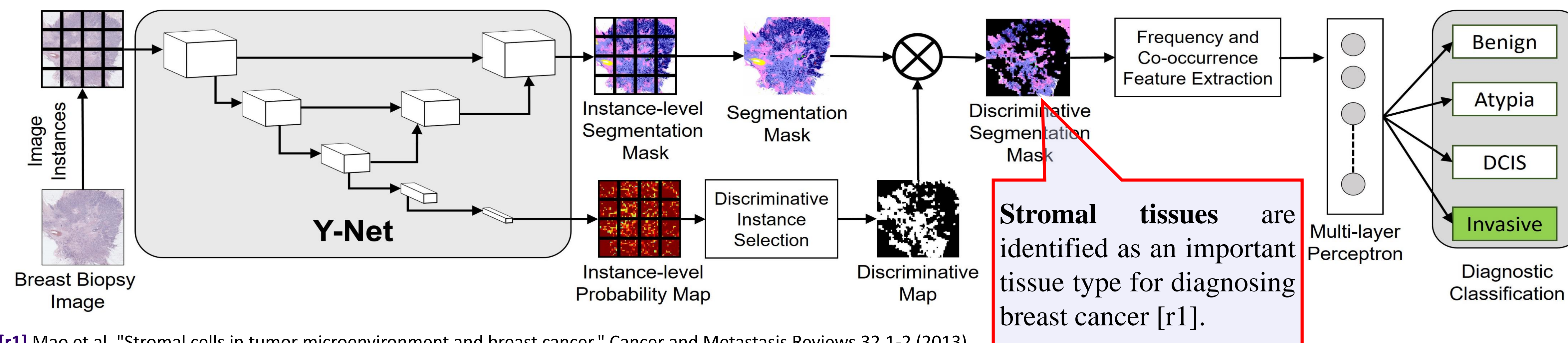- The average size of an ROI is *10,000 x 12,000* pixels.

Diagnosis	# WSIs	# ROIs (classification)	# ROIs (segmentation)
Benign	60	102	9
Atypia	80	128	22
DCIS	78	162	22
Invasive	22	36	5
Total	**240**	**428**	**58**

background
benign epi
malignant epi
normal stroma
desmoplastic str
secretion
blood
necrosis

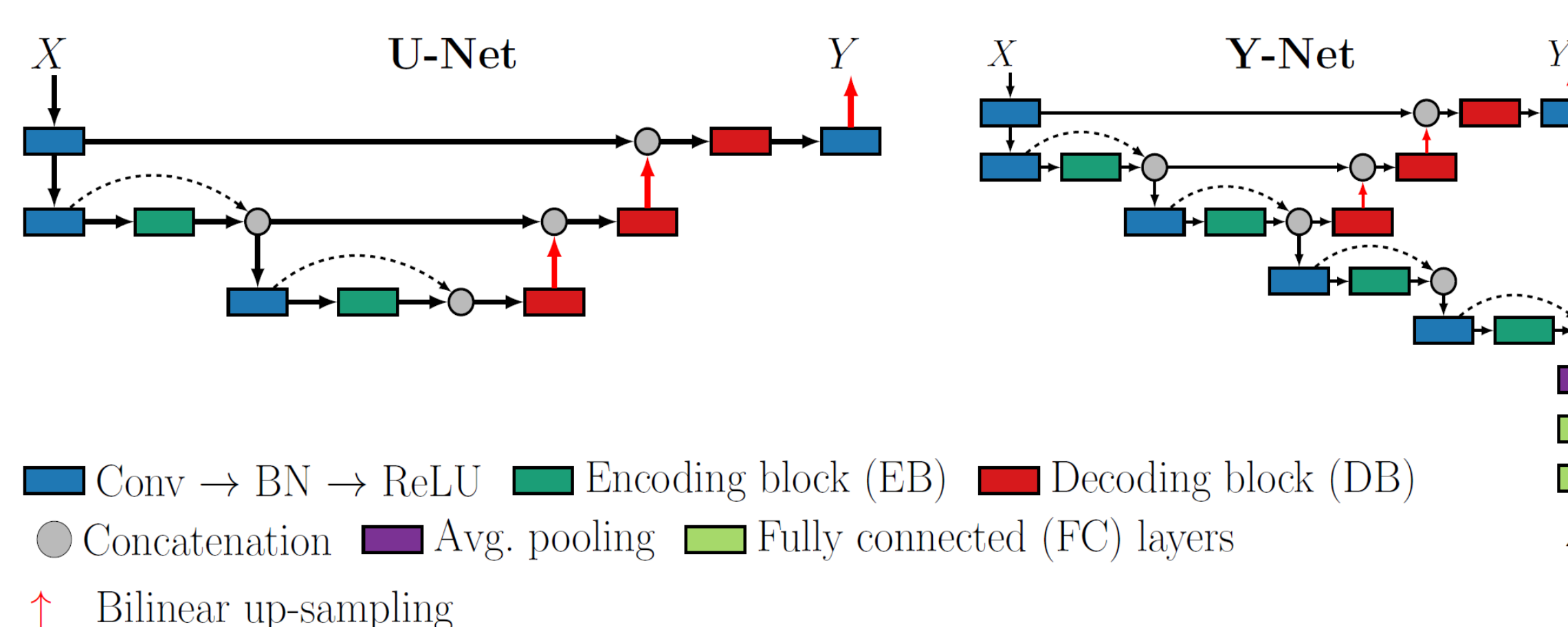## Overview of our system for detecting breast cancer

- Our system is given an ROI from a breast biopsy WSI and breaks it into instances that are fed into Y-Net.
- Y-Net produces two different outputs: **an instance-level segmentation mask and an instance-level probability map**. These outputs are then combined to produce the **discriminative segmentation mask**.
- A multi-layer perceptron then uses the frequency and co-occurrence features extracted from the final mask to predict the cancer diagnosis.
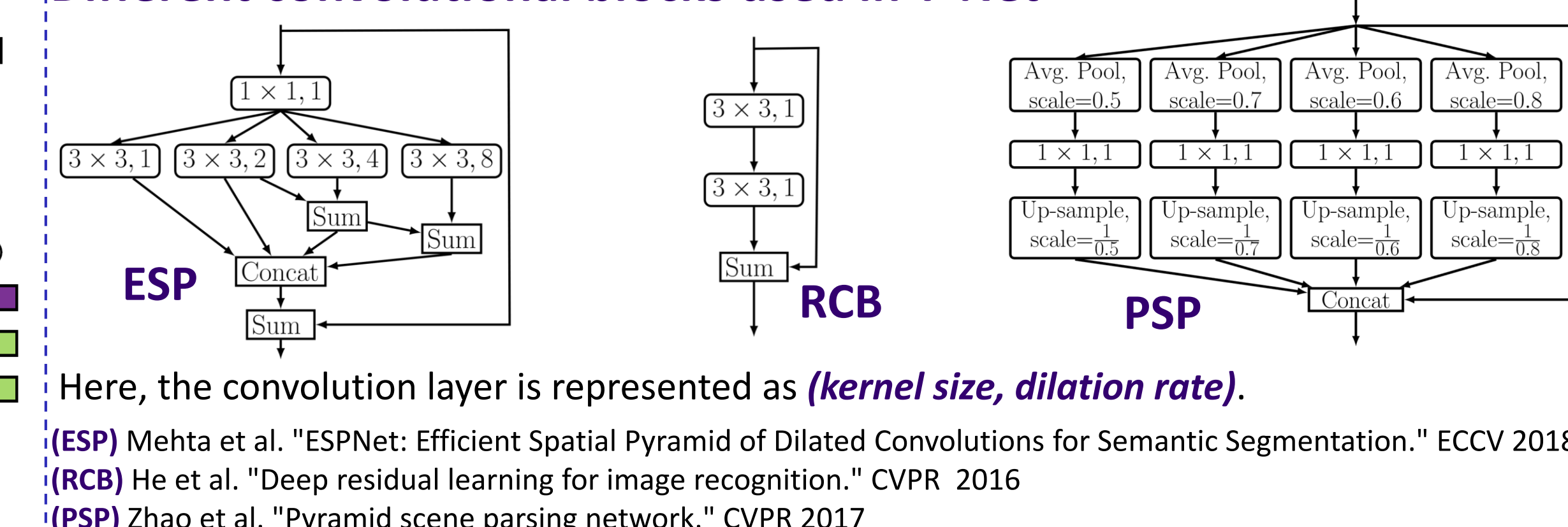

Image Instances — Breast Biopsy Image — **Y-Net** — Instance-level Segmentation Mask — Segmentation Mask — Instance-level Probability Map — Discriminative Instance Selection — Discriminative Map — Discriminative Segmentation Mask — Frequency and Co-occurrence Feature Extraction — Multi-layer Perceptron — Benign / Atypia / DCIS / Invasive — Diagnostic Classification

Stromal tissues are identified as an important tissue type for diagnosing breast cancer [r1].

[r1] Mao et al. "Stromal cells in tumor microenvironment and breast cancer." Cancer and Metastasis Reviews 32.1-2 (2013).

Y-Net architecture

- Y-Net is conceptually simple and **generalizes U-Net to joint segmentation and classification tasks**.
- U-Net outputs a single segmentation mask. Y-Net adds a second branch that outputs the classification label. The classification output is distinct from the segmentation output and requires feature extraction at low spatial resolutions.

Conv → BN → ReLU   Encoding block (EB)   Decoding block (DB)
Concatenation   Avg. pooling   Fully connected (FC) layers
↑ Bilinear up-sampling

Different convolutional blocks used in Y-Net

Here, the convolution layer is represented as *(kernel size, dilation rate)*.

(ESP) Mehta et al. "ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation." ECCV 2018
(RCB) He et al. "Deep residual learning for image recognition." CVPR 2016
(PSP) Zhao et al. "Pyramid scene parsing network." CVPR 2017

Segmentation results

- An abstract representation of encoding (EB) and decoding blocks (DB) in Y-Net allows users to more easily **explore the latent space without changing the network topology** and choose best network.
- Y-Net with ESP and PSP as encoding and decoding blocks delivered **the same segmentation accuracy** as SOTA, while learning **9.5x fewer parameters.**

EB	DB	# Params (in million)	Mean Intersection over Union
ESP	ESP	**2.25**	38.03
RCB	RCB	7.16	40.23
ESP	PSP	2.75	44.03
RCB	PSP	7.62	44.19
State-of-the-art		26.03	**44.20**

Why Y-Net?

Differentiates between relevant and irrelevant tissues automatically

- Y-Net identified stroma as more important tissue type than **blood**. This observation is consistent with the findings of pathologists.

Tissue Label   Tissue Prediction   Discriminative Map

Suppresses tissue-level misclassifications

- Tissue-wise labeled data is limited because labeling is time-consuming and requires expert pathologists. Therefore, predicted tissue-level segmentation masks are noisy and hinder classification performance.
- Y-Net suppresses tissue-level misclassifications automatically. For example, Y-Net identified incorrectly classified tissue (**secretion** is predicted as **desmoplastic stroma**) as irrelevant.

Tissue Label   Tissue Prediction   Discriminative Map

Improves diagnostic classification accuracy

- Discriminative masks produced by Y-Net **improve the accuracy by 7%** over state-of-the-art methods.

Saliency map
Segmentation mask
Discriminative mask

Accuracy (%)